DOCUMENT RESUME

ED 307 814 FL 018 024

AUTHOR Mestre, Jose P.; Royer, James M.
TITLE Cultural and Linguistic Influences on Latino Testing.
INSTITUTION Massachusetts Univ., Amherst. Scientific Reasoning Research Inst.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO SRRI-205
PUB DATE Dec 88
GRANT NSF-BNS-8511069
NOTE 61p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Bilingual Education Programs; *Cognitive Ability; *Cultural Influences; English (Second Language); Futures (of Society); *Hispanic Americans; *Language Proficiency; *Language Role; Language Tests; Listening Comprehension; Measurement Techniques; Reading Comprehension; Second Language Instruction; Spanish; Spanish Speaking
IDENTIFIERS Holyoke Public Schools MA; Sentence Verification

ABSTRACT

An overview of contemporary assessment practices often used with Latino students is provided, and it is maintained that these practices often overlook the influence of culture and linguistic proficiency on cognitive performance. A new approach, the Sentence Verification Technique (STV), is proposed as an alternative method to assess the linguistic proficiency of language minority populations. The STV, developed as a technique for measuring reading and listening comprehension, requires the student to develop one of four types of test sentences from each sentence appearing in a test passage. The four sentence types include: copying an original sentence, paraphrasing a sentence, changing the meaning of a sentence, and constructing a distractor sentence that is similar in vocabulary and syntactic structure to the sentences in the passage but unrelated in meaning. The effective application of the STV in a transitional bilingual education program in Holyoke, Massachusetts, is described, and the future needs and directions of linguistic proficiency assessment of Latino students are discussed. (DJD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED307814

SRRI #205

Cultural and Linguistic Influences on Latino Testing

Jose P. Mestre and James M. Royer

University of Massachusetts, Amherst

Scientific Reasoning Research Institute

December, 1988

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Litterer, D

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL018024

Running Head: Issues in Latino Testing

This paper was supported by a grant from the National Science Foundation, NSF #BNS-8511069.

BEST COPY AVAILABLE

Cultural and Linguistic Influences on Latino Testing

The juxtaposition of two facts foreshadow a problem looming in American education. Fact one is that Latinos are the most rapidly growing minority group with demographic trends indicating that they will be the largest minority group in this country by the turn of the century. Fact two is that Latinos score significantly below the majority population in standard assessments of both academic aptitude and academic achievement. The challenge for our instructional and assessment systems is to address the problem described in the second fact by devising approaches to accurately assess and raise the educational attainment of Latino students. Indeed, this is an important challenge since maintaining the competitive edge that is crucial to the country's economic well-being requires a highly educated workforce.

What is the cause of the academic underachievement of Latino students in the U.S.? If there were a simple answer to this question we could immediately set about making the appropriate changes in our educational system. Any reasonable attempt to formulate an answer to this question would need to draw on the interplay of socio-economic, cultural, developmental and linguistic factors. Research findings do point to linguistic proficiency as the single, most important mediator of academic achievement for Latino students (Cummins, 1981, 1982; De Avila, 1988). However, we should heed De Avila's warning that

"linguistic proficiency in English, although necessary, does not seem to be a sufficient condition [for academic achievement]" (p. 116).

Because linguistic proficiency is such an important mediator of academic achievement, the ability to accurately assess linguistic proficiency among Latino students is of paramount importance. Although this will be a focal issue of this chapter, it is important to begin with an overview of ways in which culture and linguistic proficiency can affect cognitive performance, and to illustrate how assessment practices often overlook the influence of these two factors. We then describe a new approach for assessing linguistic proficiency, called the Sentence Verification Technique (SVT), that is particularly well suited for language minority populations, and discuss how this approach has been used successfully in a transitional bilingual education program. We conclude with some thoughts on the future of linguistic assessment among Latino students.

The Interplay of Culture in Cognitive Performance

The effect of culture on cognitive performance has been documented in various studies. At one extreme is the interesting example of the Oksapmin culture of Papua, New Guinea. The Oksapmin were an isolated stone age culture until the latter part of this century when increased contact and trade with western civilization brought modifications to their culture. The

influence of western culture was perhaps manifest most clearly in the Oksapmin number system (Saxe, 1985). The Oksapmin number system consists of 27 body parts with no base structure. To count in this system, Oksapmin begin with the thumb of one hand and recite body part names moving around the upper periphery of the body. Although this system was perfectly suitable for their cultural needs (e.g., counting a set of valuables, or indicating the ordinal positions of two villages along a path), it was unsuitable for dealing with the base 10 system required for monetary exchange during trading activities. Interestingly, observations of Oksapmin children attending "bush" missionary schools revealed that they were trying to adapt and use their body part counting system to solve base 10 arithmetic problems. This example clearly illustrates how children use and adapt knowledge acquired in the home culture within an academic setting.

A similar cultural influence on mathematical problem solving can be found among the Ute Native American tribe of northeastern Utah (Leap, 1988). Leap found that Ute students often evaluated the "truth value" of a particular word problem and worked with that problem according to the findings of that evaluation. For example, when a Ute student was asked during a problem solving interview to determine how much money his brother would have to spend on gasoline if he wanted to drive his truck from the reservation to Salt Lake City, the student did not attempt to solve the problem based on the information presented in the request. Rather, he assessed the truth value of the request and

answered, "My brother does not have a pick-up truck." According to Leap, such assessments are a common part of everyday life in the reservation and are related to the Ute language, which has very precise mechanisms for evaluating the "degrees of reality" of an event. Although it could be argued that the primary link to this behavior is language and not culture, this is clearly not the case: The Ute students in the study were not fluent in the Ute language, although Ute was used in the home. Thus, despite their lack of fluency in the Ute language, Ute children adopt tribal traditions and use them in their approach to mathematical problem solving.

Among Mexican-Americans, most studies of the role of culture on achievement have focused on familism (MacCorquodale, 1988). Familism refers to the relative importance of family members in determining an individual's values, goals and orientation. Some argue that the family obstructs intellectual development because the needs of the family are supposedly placed above those of an individual family member (Grebler, Moore and Guzman, 1970; Montez, 1960). There is evidence that Mexican-Americans who are more independent of their families exhibit greater educational achievements than those who maintain closer family ties (Schwartz, 1971). However, others argue that family orientation does not interfere with either aspirations or achievement (Lopez-Lee, 1972). Research findings support this view--Anglos and Mexican-Americans do not differ in educational values and aspirations (Aiken, 1979; Espinoza, Fernandez and Dornbusch, 1977; Juarez &

Kuvlesky, 1969). Regardless of which view is correct, there does appear to be a clear difference between Anglo and Mexican-American family values in one area: Mexican-American parents are more traditional in their attitudes toward gender roles than Anglo parents are; Mexican-American girls are likely to be encouraged to pursue careers that will not interfere with their future family life (MacCorquodale, 1988).

In contrast to what one might expect given the performance differences between Anglos and Mexican-Americans in national assessments, research evidence suggests that Mexican-Americans are more likely than Anglos to do homework and that Mexican-American parents are very supportive of their children's education (MacCorquodale, 1984, 1988). These findings raise the interesting question, "To what can the differences in educational attainment between Anglos and Mexican-Americans be attributed?" MacCorquodale argues that Mexican-American parents are unable to translate their encouragement and support into concrete actions (e.g. helping their children with homework or advising their children on what courses to take in school), in part due to their limited educational background.

Although it could be argued that the examples discussed above would have an indirect effect on cognitive performance, there is evidence that culture can directly affect one of the more important instances of cognitive performance: reading comprehension. For example, one study investigated the reading comprehension of students from two distinctly different cultural

groups (Americans and Asian Indians) on two stories, one based on their own culture and one based on the foreign culture (Steffensen, Joag-Dev & Anderson, 1979). The two stories described a wedding in the U.S. and an Asian Indian wedding. After reading passages describing both weddings, subjects were able to recall considerably more material that was culturally similar to them, and also rated their understanding of the culturally familiar material higher.

Similar findings were obtained in a study of people's comprehension of baseball (Voss, Vesonder & Spilich, 1980). When passages describing the events transpiring during a baseball game were read by individuals possessing high-knowledge and low-knowledge of the game of baseball, the high-knowledge individuals were better able to recall the salient features of the baseball passage. In this example, being well-versed in the "culture" of baseball served as a clear advantage in reading comprehension. Similar findings were obtained in a study of children's comprehension of a passage about spiders (Anderson & Shifrin, 1980). It therefore appears that the presence of material that is culturally foreign to students will adversely affect performance on tests, even though the students may have mastered the skills that the test is supposedly testing.

Culturally unfamiliar material has an adverse effect not only on performance but also on learning. A study of ninth grade Latinos learning algebra (Mestre, 1988) revealed that the students were not reading the textbook as a means of supplemental

instruction. Because students did not read the textbook, the only instruction they received consisted of the in-class presentations by the teacher. Although readings from the textbook were assigned, students used the book only as a place to find problems assigned for homework. Upon closer scrutiny, it became evident that the context of the material in the textbook was largely unfamiliar to the students. When students were asked to explain the meaning of terms appearing in a typical section of the book, such as "shares of stock," "revolving charge account," "monthly payments," and "interest," it became clear that they had little idea what these terms meant. The context of the material in the textbook might be suitable for students from middle or high socioeconomic backgrounds, but it was totally unsuitable for the low socioeconomic backgrounds of the Latinos in that algebra class, a criticism that we are not the first to make (Taylor, 1978).

The Interplay of Language in Cognitive Performance

The lack of language proficiency can adversely affect cognitive performance through a variety of avenues, some more blatant than others. It is fairly easy to detect the more obvious ways in which language adversely affects cognitive performance, such as the case of a student asked to perform a task in a language with which she or he is not proficient. We will focus our discussion on the more subtle avenues through which language

proficiency affects cognitive performance, since these are not as easy to detect.

Before proceeding further it is important to define language proficiency since studies often attribute the poor performance of language minority students to lack of language proficiency without ever assessing it (De Avila, 1988); clearly these studies equate language proficiency with ethnicity. We will define a language minority student to be proficient in the second language if that student's language proficiency level is equivalent to the "average" monolingual student of the same age. We take it as self-evident that a language minority student who is not proficient in the second language, as we have defined the term, will not perform as well on a task requiring language proficiency skills as will a mainstream monolingual student. The next obvious issue to consider is how likely it is for a language minority student to be proficient.

Most bilingual programs in this country are "transitional" programs, meaning that the school system has 3 years to bring the second language proficiency level of the students to a level where they can be mainstreamed. However, there is evidence that this is not nearly enough time to make students proficient. Indications are that it takes between five and seven years for language minority students to approach language proficiency levels equivalent to their monolingual peers (Cummins, 1982). Nevertheless, often language minority students are mainstreamed before they are proficient, and many subsequently exhibit below

average performance in mainstream classrooms and on assessments of academic achievement. The reason they frequently have difficulty can be found in the relationship between language proficiency and cognitive performance.

After approximately two years of exposure to English, many language minority students display sufficient skills in English to be able to communicate quite adequately with their monolingual peers in face-to-face situations. This type of "proficiency" is frequently presumed to be sufficient to allow students to enter the mainstream curriculum and compete favorably with monolingual students, a presumption that is inaccurate. According to Cummins (1980, 1981, 1982) it is important to distinguish between the language proficiency needed for face-to-face communications and the language proficiency needed for academic work. Face-to-face communication is "context-embedded" in the sense that there are many cues to aid communication (e.g. gestures and intonation). On the other hand, much academic work takes place within a "context-reduced" situation where the cues present in face-to-face communication do not exist. For example, in attempting to read and comprehend a complex text all a student has to go on are the words in the text. Ability to communicate in context-embedded situations will not necessarily help a language minority student perform in a context-reduced academic task.

This was the situation experienced by the first author, who was mainstreamed after spending two and one-half years in a bilingual program. He was able to communicate tolerably well with

his monolingual peers, and even perform at "above average" level in school. In those cases where he did not understand the teacher's instruction or assigned work, he would turn to a neighboring student and ask for clarification. Thus, by forcing a context-reduced situation into a context-embedded situation, he was able to function quite well. However, it was in tests requiring substantial linguistic proficiency (both in-class and standardized) that problems arose. In these context-reduced assessments, the only cues available were the words on the paper. Particularly problematic were tasks where a passage was to be read followed by a series of "comprehension" questions; if several key words in either the passage or the questions were unfamiliar, there was nothing to do but guess and hope for the best.

In addition to the above unsolicited testimonial, there is research evidence of how limited English proficiency adversely affects performance on context-reduced academic tasks. One study with Latino college students revealed that some error patterns in solving math word problems were the result of language deficiencies; students were working out the problems incorrectly, but consistent with their interpretation (Mestre, 1986a). Similar evidence has led Dawe (1984) to hypothesize that, in order to perform well in cognitively demanding mathematical tasks, students must reach a threshold level of proficiency both in mathematical knowledge (i.e. mathematical concepts and how they are applied) and in the language used to express that knowledge. This language that is specific to mathematics has been termed the "mathematics

register" and is comprised of the variety of language oriented to mathematics activities, including the various linguistic forms, along with their meanings and uses (Halliday, 1975; Spanos, Rhodes, Dale & Crandall, 1988). Whereas it was previously thought that mathematics was a domain in which language proficiency played a relatively unimportant role, recent research suggests that this is not the case (Kintsch & Greeno, 1985). To be able to achieve in mathematics, students not only must be proficient in the English language, which is used throughout mathematics, but also must be proficient with the mathematics register, which defines the specific uses of the English language within mathematics.

Another example of how language interacts with cognitive performance emerged in a study of premise comprehension with both Anglo and Latino college students (Mestre, 1988). Findings from this study revealed that all students were using rules that govern the comprehension of natural discourse to interpret premises, rather than the rules of logic. For example, students were inclined to interpret the premise, "not all clerks are male" to mean "some clerks are male," rather than the appropriate interpretation, "some clerks are female" (the statement "not all clerks are male" is consistent with all clerks being female, thus it is incorrect to assume that some clerks are male). Although there were no significant differences between Anglos and Latinos in performance, an interesting difference emerged in evaluating an intervention strategy designed to improve performance.

The intervention strategy was a thirty-minute videotaped

lesson which covered the rules for parsing and interpreting premises containing different number and types of negations. Immediately following the lesson, all Anglo students reached ceiling level performance on a post test assessing their ability to interpret premises. One week following the lesson, all Anglo subjects had retained a ceiling level performance. Six months later, 93% of the Anglo subjects had retained ceiling level performance. The pattern for the Latino subjects was different--only 65% reached ceiling level performance immediately following the lesson. Among those Latinos who had reached ceiling level performance after the lesson, only 38% were able to attain ceiling level performance six months after the lesson. This disparity in learning and retention from a context-reduced, linguistically laden lesson was attributed to the disparity in English proficiency between the Anglo and Latino groups; the average SAT-Verbal score of the Anglo group was a full 170 points higher than that of the Latino group. This study illustrates both the short and long-term effects of linguistic deficiencies on learning and performance in academic situations.

Issues of Concern With Latino Assessment

The research reviewed in the previous two sections suggests that there are special considerations involved in the construction, use and interpretation of assessment instruments among Latino students, particularly when those students are in the

transition stage from first to second language proficiency. But it is equally evident that current assessment procedures are not responsive to these considerations. As Cummins states, "Most minority language students are still ... assessed ... with assessment tools and procedures that were designed only for children from the majority Anglo group," (1982, p. 1). These assessment instruments necessarily reflect the values and culture of both those who design the instruments and the mainstream population. For example, a passage in a reading comprehension test about a nature hike through the New England woods in winter may bear no overlap with the experiences of Latino students living in Puerto Rico or southern Texas. The Latino students may lack the mental schemata that facilitate interpreting and making sense out of that passage (Cabello, 1984; Steffensen, et al., 1979; Voss, et al., 1980).

Often, poor performance on these assessment instruments by bilingual students is used to reinforce myths, such as "bilingualism causes language handicaps," or "the best way for bilingual students to make progress in the second language is to eradicate the first language." There is research evidence that these myths are false (Cummins, 1981; Kessler, 1987; Leap, 1988).

Further, attempts to "patch up" assessment instruments designed for the majority population to make them suitable for Latino populations are often flawed. The most common patch-up consists of translating a test into Spanish. This practice places the student in double jeopardy--not only will the cultural values

of the test remain intact, but there is now the risk that the translation will not be adequate for the target group (Cabello, 1984); for example, differences in the level of vocabulary across the two versions of the test, and the multiplicity of translations for particular words (e.g. "kite" could be translated as "papagayo," "cometa" or "papelote" depending on the country of origin of the translator), are just two of several possible limitations in translated tests (Wilen & Sweeting, 1986).

The analysis presented thus far suggests that current assessment procedures may be appropriate for Latino students when they have acquired academic-level proficiency in English and when they have acquired sufficient familiarity with American culture, but they are inadequate procedures for assessment during the period when the students are gaining second language proficiency and are being aculturated. Given the shortcomings with current assessment procedures, what would the ideal assessment procedure look like? The literature reviewed in the previous sections suggest two guidelines. First, the ideal assessment procedure would be sensitive to cultural experiences. This suggests that the ideal procedure would be tailor-made to suit the cultural background of the target population. Mexican-American students in southern California might be assessed using one instrument, and children of Cuban background in Florida assessed with another. Second, the procedure should be sensitive to level of language proficiency. That is, the procedure should be able to trace the development of second language proficiency from the acquisition of

"face-to-face" competence to the mastery of academic level proficiency.

In the section to follow we describe an assessment procedure that attempts to be responsive to these guidelines. The first part of the section will present a general overview of the technique. This will be followed by sections that describe how the technique is responsive to the guidelines set forth, and data will be presented that addresses the validity of the procedure as a measure of progress in a transitional bilingual education program.

Overview of the Sentence Verification Technique

The Sentence Verification Technique (SVT) is a recently developed technique for measuring reading and listening comprehension that was first introduced by Royer, Hastings, and Hook (1979). The technique entails developing one of four types of test sentences from each sentence appearing in a text passage. The first type of test sentence is called an original and it is a copy of a sentence as it appeared in the passage. The second type of test sentence, called a paraphrase, is constructed by changing as many words as possible in an original sentence without altering the meaning of the sentence. The third type of test sentence is called a meaning change, and it is constructed by changing one or two words in the sentence so that the meaning of the sentence is altered. The final kind of test sentence is called a distractor

and it is a sentence that has a vocabulary level and syntactic structure that is similar to sentences in the passage and is consistent with the overall theme of the text passage, but the sentence is unrelated in meaning to any sentence that appeared in the passage.

An SVT test consists of a set of passages, each of which is followed by a set of test sentences. Each set of test sentences consists of equal numbers of each of the test sentence types. So, for example, most of the research using the SVT has used 12 sentence passages and either 12 test sentences (3 each of the test sentence types) or 16 tests sentences (4 each of the test sentence types). An examinee taking an SVT test reads or listens to the passage, and then in the absence of the text judges each of the test sentences to be "old" or "new" (the elementary school versions of SVT tests have recently started using "yes" and "no" as substitutes for old and new). Old (yes) sentences are defined as sentences that are the same as or mean the same as passage sentences (originals and paraphrases) and new (no) sentences have a different meaning than passage sentences (meaning changes and distractors). More details on the development and administration of SVT tests can be found in Royer, Greene, and Sinatra (1987), and in Royer (in press).

Theoretical Rationale for the SVT

The use of a verification technique as a measure of comprehension was shaped to a considerable degree by the theoretical assumption that comprehension is a "constructive"

process that results in a memory representation that preserves the meaning but not the form of a linguistic message. The constructivist theoretical framework (e.g., Brown, Bransford, Ferrara, & Campione, 1983; diSibio, 1982; Jenkins, 1974, 1979; Kintsch & van Dijk, 1978; Royer, 1985; Royer & Cunningham, 1981) asserts that the process of comprehension entails an interaction between context, the linguistic message, and the knowledge base of the listener or reader. This interaction results in the construction of an interpretation of a linguistic message that preserves the meaning but not the surface structure of the message. This process of forming a memory representation is thought to occur more or less simultaneously with the reception of the message (e.g., Carroll, 1972; van Dijk & Kintsch, 1983), and it is largely unconscious except in instances where processing difficulties are encountered (e.g., Kintsch & van Dijk, 1978).

Constructive theory suggests that the "product" of comprehension is a memory representation that preserves the meaning of a linguistic message. This perspective, in turn, suggests that comprehension could be measured by determining if readers or listeners had successfully established a meaning preserving memory representation of something they had read or heard. The SVT was designed to accomplish this purpose.

If readers have comprehended a text and established meaning preserving memory representations of that text, then they should be able to correctly judge that original and paraphrase test sentences have the same meaning as their memory representations,

and they should be able to correctly reject meaning change and distractor test sentences as having a meaning different than their memory representations. However, if a reader has not successfully established a meaning-preserving memory representation he or she should have great difficulty in correctly classifying the test sentences as having the same or a different meaning than a text sentence.

Scoring and Interpretation of SVT Tests

Two procedures have been used in scoring SVT tests. The first is a simple computation of proportion correct. Proportion correct can be computed for overall performance, for performance on separate passages, for performance on particular sentences within a passage, and for particular test sentence types (e.g., originals, paraphrases, etc.) within or across passages. Most of the research on the SVT has entailed calculating proportions correct, though for some purposes a more sophisticated test scoring procedure utilizing the theory of signal detection (TSD) (Swets, Tanner, & Birdsall, 1961) may yield more useful data. The scoring of SVT tests using TSD parameters may be a particularly attractive property if SVT scores were to be used as an index of the absolute comprehension of a passage. The scoring of SVT tests using TSD parameters is described in Royer (in press).

Research completed to date indicates that average readers get about 75% of the SVT items correct if the tests are based on material at grade level. With respect to listening performance, students can typically understand difficult material better when

listening than when reading, but the difference between listening and reading comprehension diminishes as the students get older.

The Reliability and Validity of SVT Tests

The reliability of SVT tests has been assessed in a number of studies (Marchant, Royer and Greene, 1988; Royer, Kulhavy, Lee, and Peterson, 1986; Royer & Hambleton, 1983; Royer, Tirre, Sinatra, and Greene, in press) involving both children and adults. In the studies involving children, 1150 students in the grades 3 through 8 were tested. The average reliability for SVT reading tests in these studies was over .9 and the reliability for listening tests was over .7. The studies involving adults assessed U.S. Air Force enlistees and college students. The reliabilities of the reading tests, which were shorter than those used in the children studies, were in the range of .6 to .7.

The research assessing the validity of the SVT as a measure of comprehension has been conducted with Messick's (1980) observation in mind that test validity is ultimately a matter of construct validity and that "...construct validation is a continuous, never-ending process developing an ever-expanding mosaic of research evidence" (Messick, 1980, p. 1019). The construct validation research on the SVT is briefly outlined below.

- The SVT is sensitive to text readability (Royer, Hastings, & Hook, 1979, Experiments 1 & 2; Royer & Hambleton, 1983; Royer, Kulhavy, Lee, & Peterson, 1986).
- The SVT is sensitive to differences in reading

skill (Royer, et al., 1979, Experiment 2; Royer, et al., 1986; Rasool & Royer, 1986; Royer & Hambleton, 1983).

- The SVT is sensitive to text characteristics (Royer, et al, 1984, Experiments 2 and 3).
- SVT performance varies as a function of working memory capacity (Lynch 1986; 1987).
- The SVT measures passage comprehension, not just sentence comprehension (Royer et al. 1984, Experiment 4).
- SVT tests have good divergent and convergent validity properties (Royer, 1986).
- The SVT can measure both listening and reading comprehension (Royer, et al., 1986; Royer, Sinatra & Schumer, in submission; Royer, Carlisle, Walbaum & Carlo, 1987; Carlo, Sinatra & Royer, 1988).
- SVT tests measure educational gain (Experiments 1 and 2 in Royer, Lynch, Hambleton & Bulgarelli, 1984; Royer et.al., 1987; Carlo et.al., 1988; Sinatra et.al., 1988).
- SVT tests predict future learning performance (Royer, Marchant, Sinatra, & Lovejoy, in submission; Royer, Abranovic & Sinatra, 1987; Marchant, Royer, & Greene, 1988).
- SVT tests have diagnostic utility (Carlisle, in submission; Royer, Sinatra, & Schumer, in submission).

Virtues of the SVT as a Measure of Language Comprehension Performance with Latino Students

In an earlier section of this paper it was suggested that assessment procedures for use with Latino students involved in

transitional bilingual education programs should be sensitive to differences in cultural background, and they should be able to assess the degree of second language competence attained by the Latino student. In this section we will consider the ways in which the SVT procedure meets these guidelines.

Developing tests that are sensitive to cultural background

From our perspective, local test development is the key to developing tests that are sensitive to cultural background. Local teachers and parents are the best judges of whether students have the cultural background that would allow them to understand a particular set of materials. Thus, teachers (possibly in conjunction with parents) could select materials that they believe to be consistent with the cultural experiences of their pupils and tests can be based on those materials.

The SVT lends itself quite well to this perspective. SVT tests can be based on virtually any text material. Moreover, it is easy to train local school personnel to develop SVT tests. This circumvents many of the problems associated with different uses of language among speakers of a language who have varying backgrounds.

As an instance of local SVT test development, the second author of this paper is involved in a research project on the Island of Grenada that is assessing the impact of Computer Assisted Instruction in a developing country. This evaluation effort included developing reading comprehension tests for grades 2 through 7 suitable for use in Grenada. The national language in

Grenada is English, but there are striking differences between English as spoken and used in the U.S., and English as spoken and used in Grenada. The process of test development involved training Grenadian teachers to develop SVT tests and then having them develop tests based on materials selected from local sources. The tests developed by the teachers have been administered twice, and all of the evidence indicates the tests are valid indicators of progress in Grenadian schools.

The Grenadian example described above, and the bilingual assessment study to be described later in this chapter, demonstrate how the SVT can be responsive to cultural background issues. The tests can be based on materials that local personnel judge to be consistent with the background experiences of the students, and the tests can be developed by people having linguistic and cultural experiences that are the same as the target population to be assessed. These procedures should result in culturally fair tests.

Developing tests that assess degree of second language competence

Earlier in this paper a distinction was made between face-to face competence in a second language and academic competence in that language (Cummins, 1980; 1981; 1982). It was suggested that accurate decisions about which of these levels of competence had been attained was critical to the correct placement of Latino students. Those who have attained academic competence in English may benefit most from placement in mainstream classrooms. However, those who have only mastered English at the face-to-face

competence level may undergo unnecessary educational hardship through placement in mainstream classrooms.

SVT tests provide a means of assessing the degree of language competence attained by students in the process of learning English. This is accomplished by administering both listening and reading comprehension tests. In our SVT studies with Latino students we have routinely administered listening and reading tests in both English and Spanish. The purpose of administering listening and reading tests in both languages is to attain a comprehensive portrait of a student's overall language ability. The listening and reading tests in Spanish provide indices of native language comprehension ability (performance on the listening tests) and reading comprehension ability. The performance on the English listening tests provide indices of face-to-face linguistic competence, and performance on the reading tests provide indices of academic competence in English. In our studies we have used audio tape recorders as a means of administering the listening tests. But if one wanted to increase the extent of "context-cues" present in the assessment session, the tests could be administered by having someone read the passages and tests.

This section has indicated how SVT tests can be responsive to some of the needs associated with assessing Latino students in the process of learning English. The next section will examine evidence that the tests are in fact fulfilling these needs.

Using SVT Tests to Track the Educational Progress of Students Enrolled in Bilingual Education Programs

This section will report the results of studies completed in the Holyoke, Massachusetts, Public School System that evaluated the validity of SVT tests as measures of educational progress for students enrolled in a transitional bilingual education (TBE) program. The procedural details of the study will be described very briefly in this chapter. The reader interested in more detail on procedural matters and a more comprehensive description of the research results can find them in Carlo, Sinatra, & Royer, 1988; Royer & Carlo, 1988; and Royer, Carlisle, Walbaum, and Carlo, 1988.

The Holyoke, Massachusetts, Public School System enrolls approximately 6,500 students, about 20% (1400) of whom are enrolled in TBE programs. The large majority of the TBE students are native speakers of Spanish and virtually all of them are Puerto Rican. The school system has two types of TBE programs. The first is a traditional model involving six steps (Levels I,IIA, IIB, IIC, III, mainstream) and the second is a "two-way" model in which TBE students spend considerable in-class time with native English speaking students.

Students with little or no competence in English are placed in Level I of the TBE program where they receive all of their content instruction in their native language, combined with English as a Second Language (ESL) instruction. As they acquire

competence in English, subject matter instruction in English is phased in beginning with mathematics, then science, then social studies, and finally reading. The phasing in process in mathematics, science and social studies occurs during Level II (A,B, and C) of the program, and English reading instruction is encountered in Level III. Students are mainstreamed after the school systems judges them to be competent in English.

When Spanish speaking students enter the system they are tested with the Bilingual Syntax Measure (BSM) and interviewed. Evaluations of student progress are conducted twice a year. The criteria for exit from a TBE classroom to a mainstream class is a satisfactory score on the BSM and a grade equivalent score on the English form of the California Achievement Test (CAT) that is at least equivalent to the students current grade placement. This latter requirement has been very difficult for many of the Spanish speaking students to achieve. Moreover, many of the native speakers of English in the school system do not meet the requirement of grade level performance on the CAT.

Methodology

The test development phase of the studies began with having teachers from the school system select reading textbooks that could serve as the basis for SVT tests in English and Spanish. The books selected by the teachers were either in use in the system or they were judged to be very representative of the type of reading material both the TBE and mainstream students would be likely to encounter in their classrooms.

The research team then selected passages from the textbooks and edited those passages so that they were 12 sentences in length, and they were "coherent" in that they had a beginning, middle, and end. The coherence editing was undertaken to avoid, as much as possible, the sense that the passages were excerpts taken from a larger text.

The next step involved developing SVT tests based on the passages. The tests were developed in accordance with the procedures described in Royer, Greene, & Sinatra (1987), and Royer (in press). The Spanish tests were developed by a graduate student who is fluent in both English and Spanish and who is a native of Puerto Rico. The English tests were developed by graduate students who are native speakers of English.

The listening and reading tests were developed using passages drawn from the same text source. When selecting passages for inclusion in the tests two passages were selected that were judged to be parallel in difficulty. One of these passages then became part of the reading test and the second passage was included in the listening test. Tests for a particular grade level were constructed by "bracketing" a grade level. That is, the tests for a given grade consisted of passages that were thought to be easier than the reading competence of the average reader, passages that were thought to be equal to the reading skill of the average reader, and passages that were thought to be more difficult than the reading skill of the average reader.

After the tests were developed they were returned to the

teachers for review and criticism. Following this review, final changes in the tests were made, and the listening tests in both Spanish and English were recorded on audio tape by the bilingual graduate student and the reading tests were prepared in booklet form.

The SVT tests have been administered to two student cohorts. The first cohort consisted originally of 115 fifth grade TBE students who were administered SVT tests in February and May, 1987, and in June 1988. The second cohort consisted originally of 120 students enrolled in the TBE program and to 260 students enrolled in mainstream classrooms. The second cohort, which consisted of students in grades 3,4,5,and 6, was administered SVT tests in December 1987, and the tests were re-administered in June, 1988. Mainstream students were given only the English listening and reading tests. The data reported here come primarily from the second cohort. However, there are several issues that will be examined using data from the first cohort.

Collection of ancillary information. The school system completed an information questionnaire for every student participating in the studies. This questionnaire asked for: 1) student age, 2) date of entry into school system, 3) Lau category, 4) TBE Level at the time the SVT tests were administered, 5) grade level of currently assigned reading book, 6) available standardized test scores, and 7) language spoken at home.

In addition to the information provided by the school system, the teacher of every student participating in the study rated each

of his or her students on their listening and reading comprehension. Teachers of TBE students rated their students on listening and reading comprehension proficiency in both Spanish and English. Mainstream teachers rated their students only in English.

Indices of Test Validity

The purpose of the study was to assess whether SVT tests were valid indicators of educational progress for students enrolled in a bilingual education program. Evidence consistent with the conclusion that the tests were valid indicators of educational progress would be present if the results followed the patterns described below.

1) Performance of the TBE students on the English tests should vary in accordance with TBE Level. That is, Level I students should score lower than Level II students, Level II students should score lower than Level III students, and Level III students should score lower than native speakers of Spanish who are in mainstream classrooms.

2) Performance of both TBE and mainstream students should vary on the reading tests as a function of level of assigned reading book. For instance, if one TBE student in the 4th grade is reading in a Spanish book at the 3.0 grade level, and a second student is reading in a 4.0 book, the second student should receive a higher score on the Spanish reading SVT test than the first student.

3) Performance of both the TBE and the mainstream students

should vary in accordance with teacher ratings of competence. As an instance, students rated as being highly competent in English reading should score higher on the English reading SVT tests than students rated at a moderate degree of competence, and students with moderate ratings should score higher that students with low ratings.

There are two other issues that can be examined in the context of the research effort. Both of these issues provide part of the underlying rationale for the beneficial effect of bilingual education programs. The first issue is the assumption that listening competence precedes reading competence. This assumption underlies the decision to provide ESL instruction prior to beginning systematic instruction in reading in English.

The second assumption is that academic skills acquired in one language will transfer to a second language when that language is acquired. Below are formal statements of these assumptions in terms of expectations about SVT performance.

4) TBE students will have better scores on English listening tests than they will have on English reading tests during the time they are acquiring competence in English.

5) Performance on the Spanish reading tests will be predictive of future skill on the English reading tests.

Results

SVI performance on the English tests as a function of TBE level. The Holyoke school system advances children through levels of the TBE program based on perceived increases in competence in

English. Therefore, if the SVT tests in English were valid indicators of educational progress for the TBE students it should be the case that SVT performance should vary in accordance with TBE level. Figure 1 presents the results corresponding to this expectation. The data in the figure are average proportion correct on the SVT tests. The data is summed over the grade level of the students for ease of presentation, but graphs drawn separately for each of the grade levels depict essentially the same pattern as that presented in Figure 1.

Insert Figure 1 About Here

The data in Figure 1 show the performance of the TBE students on the English SVT listening and reading tests administered in December, 1987 (labeled ENG LIS I and ENG RDG I in the graph), and May, 1988 (LIS and RDG II in the graph). In addition to showing data for TBE students at each level of the TBE program (the A, B, and C divisions in level 2 of the TBE program have been collapsed), the graph also indicates the performance of Latino students in mainstream classrooms on the same tests, and it indicates the performance of native speakers of English on the tests. The Latino students in mainstream classrooms are students whom the teachers list as having Spanish as the language spoken at home. Many of these students were undoubtedly graduates of the TBE program, but it was not possible to sort out former TBE students from Spanish speaking students who had never entered the

TBE program.

As can be seen in the figure, there is a clear relationship between placement in the educational program and performance on the SVT tests. Performance on both the listening and the reading English SVT tests increases as a function of advancement through the TBE program. It is also of interest that the mainstream Spanish speaking students are performing at approximately the same level as are the native English speaking students.

Another result of interest is the fact that listening performance is consistently superior to reading performance. This result is congruent with the previously mentioned hypothesis that context cues associated with speech assist the understanding of linguistic material. Moreover, it is consistent with the educational goal of attempting to develop oral understanding of English through ESL classes prior to beginning formal reading instruction in English.

The only result in the graph that is somewhat puzzling is the decline in reading performance from the first to the second testing occasion for the students in level three of the TBE program. This is curious given that every other group in the study showed a gain in performance from the first to the second testing occasion. One reason that this might have happened is that the bulk of these students were in one fifth grade classroom and the examiner reported that these students were particularly unruly on the day the tests were administered.

The data presented in Figure 1 reports performance on the

English tests. Similar analyses of performance on the Spanish tests have been conducted and they show no relationship between test performance and TBE level.

Performance on the SVT tests as a function of teacher assessments of listening and reading competence. The teachers of the students who participated in the study were asked to rate every student with respect to their listening and reading comprehension skills. These ratings were made on a 1 to 9 scale with the ratings of 1 and 9 on the scale being anchored by the very best student in the class (rating of 9) and the very worst student in the class (rating of 1). After the teachers had made their ratings the scales were collapsed to form three categories: low (ratings 1-3), medium (ratings 4-6), and high (ratings 7-9). Students in the TBE program were rated on both English and Spanish skills, whereas mainstream students were rated only on English skills.

If the SVT tests were valid indices of educational progress, performance on the listening and reading tests should vary in accordance with teacher judgments of competence. Figure 2 presents the data from the Spanish listening and reading tests. The data in the figure is average performance on the tests summed over grade. Similar graphs drawn for each grade separately show patterns much the same as the averaged data presented in Figure 2.

Insert Figure 2 About Here

As can be seen in the figure, performance on both the listening and the reading tests increases as a function of teacher judgments of competence. Graphs similar to the one presented in Figure 2 have been plotted for mainstream students and for TBE students when rated on English skills and when tested with English tests. The data for mainstream students is even more striking than that for TBE students. That is, there is a clear correspondence between performance on the tests and teacher judgments of competence.

The graph depicting the relationship between the TBE students' English SVT performance and teacher judgments were not nearly as orderly. The likely reason for this is that the TBE classrooms were segregated to some extent by TBE level. That is, some classrooms had a preponderance of students at upper levels of the TBE program whereas others contained mostly students at lower TBE levels. This resulted in teachers making very fine competence discriminations. In some cases teacher were judging degrees of competence on a 1 to 9 scale between students having virtually no English competence. In other cases teachers made discriminations on the same scale between students judged to be near mainstreaming. The net result was little relationship between the teacher judgments and SVT test performance.

Another interesting aspect of the data presented in Figure 2 is that there is a greater correspondence between teacher ratings of reading competence and SVT reading performance than there is ratings of listening competence and SVT listening performance.

This data parallels other data that we have collected with mainstream students. Teachers seem to be better judges of who are their better readers in their classrooms than they are judges of who are the better listeners. This probably reflects the generally greater emphasis placed on reading as an educational goal.

SVT performance and reading book level. One of the items of information collected in the study was the level of the reading book in which the students were currently working. If SVT reading tests were accurately measuring reading skill it should be the case that test performance varied in accordance with the level of the assigned reading book. The data on reading book level does not lend itself to summarization over grades because of variations in differing numbers of reading book levels represented at each grade. Given this, the data for only grade 6 will be presented. The graphs for the other grades show the same relationship between performance on the tests and reading book level. The data in Figure 3 shows the performance of grade 6 TBE students on the Spanish SVT tests. The reading book levels are in grade units (i.e., 4.0 is a beginning 4th grade level textbook) and are the levels of Spanish reading books in which the students are working. Very few of the TBE students were receiving reading instruction in English books.

Insert Figure 3 About Here

The figure shows a clear relationship between performance on the SVT reading tests and the level of the reading book in which the students are currently working. It is also interesting to note a relationship between the level of the reading book the students are working in and their performance on the Spanish listening tests. This result is again similar to that found in other studies (e.g., Royer, Sinatra, and Schumer, in submission) where listening ability of good readers typically exceeds the listening ability of poor readers.

Graphs similar to the one presented in Figure 3 have been examined for mainstream students. They look very similar to Figure 3 in that there was a clear relationship between the level of the English reading book the mainstream students were working in and their performance on the English listening and reading tests.

Does linguistic competence transfer? One of the most fundamental assumptions underlying transitional bilingual education programs is that educational skills acquired in a native language will transfer to a developing second language (e.g., Cummins 1983, 1984; Hakuta, 1986). Despite the importance of this issue, very little empirical research has been done in this area as noted by Hakuta in the following statement:

> "What is remarkable about the issue of transfer of skills is that despite its fundamental importance, almost no empirical studies have been conducted to understand the characteristics or even to demonstrate the existence of

transfer of skills" (1986, p. 218).

The second author can verify Hakuta's observation. A very thorough search of the literature has not turned up a single empirical demonstration of educational skills acquired in a native language transferring to skill in a second language.

The data acquired from the students tested in February and May 1987, and in June, 1988 can be used to evaluate the transfer issue. One-hundred fifteen fifth grade TBE students were tested in February, 1987. Unfortunately, only 49 of the original 115 students were available for testing in June, 1988, and many of these students did not have a complete data record. Table 1 presents the results of pairwise correlational analyses which were computed using all available data. This means that the N contributing to each correlation will vary. As can be seen at the bottom of the table, the smallest N contributing to a correlation was 29.

Insert Table 1 About Here

The most interesting data in the table involves the relationships between the other test scores and performance on the English listening and reading tests administered in June, 1988 (Eng Lis 3 and Eng Rdg 3 in the table). As can be seen in the table, the only significant predictor of listening 3 SVT performance was listening 2 SVT performance, and that the next best predictor was listening 1 SVT performance. This means that

the best predictor of the TBE students' English listening competence in June, 1988 was their English listening competence 12 months before, and the next best predictor was their English listening competence 17 months before.

The predictions of listening performance can be contrasted to the predictions of reading performance. As the table shows, the best predictor of English reading performance was Spanish reading performance the year before. Spanish reading performance was a better predictor than previous English reading performance.

These data must be interpreted cautiously given the small number of students involved and given the magnitude of the relationships as illustrated by the relatively low correlational coefficients. But the data certainly suggests that the English listening competence that students acquire in and out of school translates into increasing ability to understand English, and they suggest that reading skill acquired in Spanish transfers to reading skills in English as the students improved their competence in English. The data collected in December, 1987 and June 1988 from the much larger student cohort will be valuable in assessing the transfer issue in the future.

Summary of the Empirical Studies

In the early part of this section three checks on the validity of SVT tests as indices of progress in bilingual education programs were suggested. The first check was that performance on the English versions of the tests should vary in accordance with placement of students in the TBE program. The

second check was that performance on the tests should vary in accordance with teacher judgments of competence. And the third check was that performance on the reading tests should vary in accordance with the grade level of the assigned reading book.

Performance on the SVT tests were congruent with all three of these expectations. The TBE students improved in performance on the English listening and reading tests as they advanced through levels of the TBE program, performance on the tests varied in accordance with teacher expectations of listening and reading competence, and performance on reading tests were better for students assigned upper level reading books.

In addition to providing evidence regarding the validity of SVT tests in a bilingual context, the studies that have been conducted thus far provide some support for assumptions underlying TBE programs. Specifically, the results indicated that early competence in listening and understanding English was related to English listening competence a year later, and that reading competence in Spanish was predictive of subsequent reading competence in English. These results are consistent with the practices of providing ESL instruction as part of a TBE program, and with providing reading instruction in the native language during the period of acquiring competence in the second language.

Future Directions in SVT Assessment

Work that is currently underway will expand the scope of SVT assessment in three ways. First, we are examining whether SVT assessment is useful in content areas. In a pilot study using 7th

grade TBE and mainstream students, SVT tests have been developed based on materials drawn from science and social studies text sources. If this study demonstrates that SVT tests can be used to assess the degree to which students can read and understand content area materials we will expand the scope of this work to encompass more grade levels and content areas (e.g., mathematics). This research will build upon previous research which has shown that SVT test performance on content materials can be used to predict learning performance in that content area (Royer, Abranovic & Sinatra, 1987; Royer, Marchant, Sinatra & Lovejoy, in submission).

A second future concern is to develop better ways to train local school personnel to develop their own SVT tests. One approach to SVT training that we are currently working on involves developing computer-based training programs. Two programs are under development. The first is named the SVT Test Maker (Walczyk & Royer, in submission). The SVT Test Maker is a program that automates many of the details associated with SVT test development. The test developer types in the text that is to serve as the basis for the SVT test, and the program parses the text and then asks the developer several question about how the test should be arranged. After answering these questions the program presents the developer with each sentence in the original passage and indicates that a test sentence of a particular type should be developed based on the presented sentence. The developer then types in the test sentence. After each test

sentence has been developed, the program automatically formats the test, and prints it out, complete with instructions, in a form that is ready to be reproduced and administered. The SVT Test Maker currently exists only in an English version, but after further refinements our intent is to develop it in a Spanish version.

The second computer program will be called the SVT Trainer. This program will be a tutorial that will train local school personnel to develop SVT tests. It will include an introduction to SVT testing, a brief description of prior research, and considerable practice with appropriate feedback on the development of SVT test items. The trainee will also receive instruction on constructing SVT tests, administering SVT tests, scoring SVT tests, and interpreting test results.

The final SVT project that is envisioned for the future is a computer based reading diagnostic system that will include SVT listening and reading tests. This diagnostic system is designed to assist in diagnosing reading difficulties in students who do not seem to benefit from normal reading instruction. As with the other systems, the diagnostic system could be developed in Spanish as well as English.

Concluding Remarks

We view Latino testing as being divided into two separable concerns. The first concern is that tests used to assess Latino students be sensitive to cultural and linguistic influences during the time the student is acquiring second language proficiency and

being aculturated. The second concern is that tests be developed that are not systematically biased against Latino students who have developed second language proficiency and who have achieved a degree of acculturation.

The first of the above concerns is critical to the effective education of Latino students and we believe that it can only be addressed at the local level. The introduction to this chapter reviewed evidence that culture and language can influence cognitive performance in general and test performance in particular. In our view there is no way that a single test can be truly fair to Latino students having differing cultural experiences and linguistic traditions. Our solution to this dilemma is to argue for tests that are tailor-made to fit the cultural and linguistic experiences of the target population.

Until recently, the argument for local based assessment would have been a hopeless expectation. Teacher-made tests are notorious for their poor reliability and validity, and local school systems do not have the resources to meet the enormous costs of developing reliable and valid tests using traditional psychometric procedures. We have presented evidence that SVT tests provide one means of meeting the need for local based assessment at a cost that local school systems can easily afford. We are also confident that once the idea that quality local assessment is possible, other procedures will be developed to meet this need.

We also believe that progress is being made on the

development of standardized tests that are not systematically biased against minority students. For example, one area in which test makers are making considerable progress is in test bias detection. Educational Testing Service has recently implemented a major effort for monitoring test bias through a technique called "differential item functioning." This technique compares the performance (at the test item level) of "target groups" (women, Blacks, Latinos, Native Americans and Asian Americans) against that of the majority group. If either a target group or the majority group displays differentially lower performance on an item, that item undergoes close scrutiny; following the scrutiny, the item may be deemed inappropriate and not counted in the test.

In addition, recent positive steps have been taken by test makers to help Latino students in the areas of test familiarity and test-taking strategies. Standardized tests, such as the SAT, have very clear formats. Familiarity with the format of the various sections will reduce both the anxiety that a student will experience in taking the test and the time that student must spend figuring out what a particular test section is asking. For example, the SAT-math has a section called "quantitative comparisons" where the test taker must compare two quantities and decide if: a) the quantity in the first column is less than the quantity in the second column, b) the quantity in the first column is greater than the quantity in the second column, c) the two quantities are equal, or 4) there is not enough information to determine the relative sizes of the two quantities. A student

walking into the SAT without being extremely familiar with the four answer categories for these items will be at a clear disadvantage since that student will have to spend considerable time looking back at the four options before responding to the items.

Further, there are sound test taking strategies that can help any student's performance in standardized multiple choice tests. For example, students who are familiar with the optimal "guessing strategies" in those types of tests that correct for possible guessing by subtracting a percentage of the incorrect answers from the correct answers will likely show a higher performance than students who are unaware of these strategies. One such strategy consists of instructing students in the advantage of guessing in those cases where the possible choices have been narrowed down to two or three from a field of five, while a complementary strategy consists of explaining to students why wild guessing will not help in the slightest. Finally, a wise strategy in timed multiple choice tests is not to spend too much time on any single item--a strategy which if ignored can increase the speededness of a test.

As in the case of monitoring possible item bias, Educational Testing Service and the College Board are taking an initiative in helping Latino students in both of these areas. A recent Educational Testing Service-College Board publication entitled "Preparing for the PSAT NMSQT for Latino High School Students" addresses these two specific issues for the Preliminary Scholastic

Aptitude Test.

Despite these encouraging trends, there is a need for more research focusing on testing issues with minority populations, as well as more cooperation and communication between researchers and test makers. For example, two areas where more research should be conducted with minority populations are test speededness and test validity. Because language minority students often read slower than majority students, they are likely to reach fewer items in a timed test and thereby exhibit a poor performance (Mestre, 1986b); it is therefore important to have a thorough understanding of the possible adverse effect of speededness on test performance for language minority students.

Test validity, which refers to whether or not a particular test is in fact an accurate measure of what it purports to measure, can be an unwieldy research area. It is conceivable that a test could be a valid instrument for a language minority group yet be insensitive in the cultural/linguistic dimension, or conversely that a test could be designed to be culturally and linguistically sensitive for a language minority group yet not be valid for that group. What is clear from existing research on validity is that tests which appear to be good predictors of future cognitive performance for majority students are not as good for predicting performance for minority students (Dalton, 1974; Hedges & Majer, 1976; Houston, 1980; McCormack, 1983; Mestre, 1981). These studies suggest that caution must be exercised in how we interpret test scores among minority populations.

The impression with which we hope to leave the reader is that despite the enormous problems that remain in the area of testing language minority populations, progress is being made. We believe testing procedures now exist that will allow local school personnel to develop tests that are sensitive to the cultural and linguistic backgrounds of their students. We also believe that standardized tests developers are very concerned about issues of test bias and test validity as it relates to the assessment of minority populations, and that steps are being taken to reduce the possibility that standardized tests are biased and invalid measures of the abilities of minorities.

References

Aiken, L.R. (1979). Educational values of Anglo-American and Mexican-American college students. The Journal of Psychology, 102, 317-321.

Anderson, R.C. & Shifrin, Z. (1980). The meaning of words in context. In R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., Bransford, J. D., Ferraro, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), Carmichaels Manual of Child Psychology (Vol. 1). New York: John Wiley.

Cabello, B. (1984). Cultural interference in reading comprehension: An alternative explanation. Bilingual Review, 11, 12-20.

Carlisle, J. F. (in submission) Diagnostic assessment of comprehension: Profiles in listening and reading.

Carlo, M. S., Sinatra, G. M., & Royer, J. M. (1988). Using the Sentence Verification Technique to measure transfer of comprehension skills from native to second language. Paper accepted for presentation at the 1989 Annual Meeting of the American Educational Research Association, San Francisco, California.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In J. B. Carroll & R. O. Freedle (Eds.),

Language comprehension and the acquisition of knowledge. Washington, DC: Winston & Sons.

Cummins, J. (1980). Psychological assessment of immigrant children: Logic or intuition? Journal of Multilingual and Multicultural Development, 1, 97-111.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In Schooling and Language Minority Students: A Theoretical Framework, compiled by the California State Department of Education, Los Angeles: National Dissemination and Assessment Center.

Cummins, J. (1982, February). Tests, achievement, and bilingual students. Focus, 1-8. Rosslyn VA: National Clearinghouse for Bilingual Education.

Dawe, L. (1984, August). A theoretical framework for the study of the effects of bilingualism on mathematics teaching and learning. Paper presented at the Fifth International Congress on Mathematical Education, Adelaide, Australia.

De Avila, E.A. (1988). Bilingualism, cognitive function, and language minority group membership. In R. Cocking & J. Mestre (Eds.), Linguistic and Cultural Influences on Learning Mathematics (pp. 101-121). Hillsdale, NJ: Lawrence Erlbaum Associates.

diSibio, M. (1982). Memory for connected discourse: A constructivist view. Review of Educational Research, 52, 149-174.

Espinoza, R.W., Fernandez, C. & Dornbush, S.M. (1977). Chicano perceptions of high school and Chicano performance. Atzlan, 8, 133-155.

Grebler, L., Moore, J.W. & Guzman, R.C. (1973). The family: Variations in time and space. In I.L. Duran & H.R. Bernard (Eds.), Introduction to Chicano Studies (pp. 309-331). New York, NY: Macmillan. (Originally published in 1970).

Hakuta, K. (1986). Mirror of Language: The Debate on Bilingualism. New York: Basic Books, Inc.

Halliday, M.A.K. (1975). Some aspects of sociolinguistics. In E. Jacobson (Ed.), Interactions Between Linguistics and Mathematics Education (pp. 64-73). Final Report on the Symposium Sponsored by UNESCO, CEDO and ICMI, Nairobi, Kenya, Sept. 1-11, 1974. (UNESCO Report No. ED-74/CONF. 808. Paris: UNESCO)

Hedges, L.V. & Majer, K. (1976). An attempt to improve prediction of college success of minority students by adjusting for high school characteristics. Educational and Psychological Measurement, 36, 953-957.

Houston, L.N. (1980). Predicting academic achievement among specially-admitted Black female college students. Educational and Psychological Measurement, 40, 1189-1195.

Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it! American Psychologist, 29, 785-795.

Jenkins, J. J. (1979). Four points to remember. A tetrahedral model of memory experiments. In L. S. Cernak & I. M. Craik

(Eds.), Levels of processing in human memory. Hillsdale, NJ: Erlbaum.

Juarez, R.Z., & Kuvlesly, W.P. (1969). Ethnic group identity and orientations toward educational attainment: A comparison of Mexican-American and Anglo boys. Washington, D.C.: National Institute of Education. (ERIC Document Reproduction Service No. ED 023 497).

Kessler, C. (1987). Language minority children's linguistic and cognitive creativity. Journal of Multilingual and Multicultural Development, 8, 173-186.

Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92, 109-129.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Leap, W. L. (1988). Assumptions and strategies guiding mathematics problem solving by Ute Indian students. In R. Cocking & J. Mestre (Eds.), Linguistic and Cultural Influences on Learning Mathematics (pp. 161-186). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lopez-Lee, D. (1972). The academic performance and attitudes among Chicanos and Anglos in college. The Journal of Mexican-American Studies, 1, 201-222.

Lynch, D. J. (1986). Reading comprehension performance as a function of individual differences in working memory for texts

of varying reading difficulty. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Lynch, D. J. (1987). Reading comprehension under listening, silent, and round robin reading conditions as a function of text difficulty and working memory. Paper presented at the Annual Meeting of the American Educational Research Association, Washington D. C.

Marchant, H. G., Royer, J. M., & Greene, B. A. (1988). Superior reliability and validity for a new form of the Sentence Verification Technique. Educational and Psychological Measurement, 48, 827-834.

MacCorquodale, P. (1988). Mexican-American women and mathematics: Participation, Aspirations, and Achievement. In R. Cocking & J. Mestre (Eds.), Linguistic and Cultural Influences on Learning Mathematics (pp. 137-160). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacCorquodale, P. (1984). Social influences on the participation of Mexican-American women in science (Final report). Washington, D.C.: National Institute of Education. (ERIC Document Reproduction Service No. ED 234 991).

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Mestre, J.P. (1981). Predicting academic achievement among bilingual Hispanic college technical students. Educational and Psychological Measurement, 41,

1255-1264.

Mestre, J.P. (1986a). Teaching problem solving strategies to bilingual students: What do research results tell us? International Journal of Mathematics Education in Science and Technology, 17, 393-401.

Mestre, J.P. (1986b). The Latino science and engineering student: Recent research findings. In M. Olivas (Ed.) Latino College Students (pp. 157-192), New York, NY: Teachers College Press.

Mestre, J.P. (1988). The role of language comprehension in mathematics and problem solving mathematics and problem solving. In R. Cocking & J. Mestre (Eds.), Linguistic and Cultural Influences on Learning Mathematics (pp. 201-220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Montez, P. (1960). Some difference in factors related to educational achievement of two Mexican-American groups. San Francisco, CA: R & E Research Associates.

Rasool, J. M., & Royer, J. M. (1986). Assessment of reading comprehension using the Sentence Verification Technique: Evidence from narrative and descriptive texts. Journal of Educational Research, 79, 180-184.

Royer, J. M. (in press). The Sentence Verification Technique: A new direction in the assessment of reading comprehension. In J. Algina and S. Legg (Eds), Cognitive Assessment of Language, Mathematics and Science Outcomes. Norwood, N. J.: Ablex.

Royer, J. M. (1985). Reading from the perspective of a biological metaphor. Contemporary Educational Psychology, 10, 150-200.

Royer, J. M. (1986). The Sentence Verification Technique as a measure of comprehension: Validity, reliability, and practicality. Unpublished manuscript.

Royer, J, M., & Carlo, M. S. (1988). Final report on the Holyoke Public School System Language Comprehension Assessment Project. Department of Psychology, University of Massachusetts.

Royer, J. M., & Cunningham, D. J. (1981). On the theory and measurement of reading comprehension. Contemporary Educational Psychology, 6, 187-216.

Royer, J. M., & Hambleton, R. K. (1983). Normative study of 50 reading comprehension passages that use the Sentence Verification Technique. Unpublished study.

Royer, J. M., Abranovic, W. A., & Sinatra, G. M. (1987). Using entering reading performance as a predictor of course performance in college classes. Journal of Educational Psychology, 79, 19-26.

Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. Journal of Reading Behavior, 11, 355-363.

Royer, J. M., Greene, B. A., & Sinatra, G. M. (1987). The Sentence Verification Technique: A practical procedure teachers can use to develop their own reading and listening

comprehension tests. Journal of Reading, 30, 414-423.

Royer, J. M., Sinatra, G. M., & Schumer, H. (in submission). Individual patterns in the development of listening and reading comprehension.

Royer, J. M., Carlisle, J., Walbaum, S., & Carlo, M. S. (1988). A new procedure for measuring progress in a Transitional Bilingual Education program. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, La.

Royer, J. M., Kulhavy, R. W., Lee, J. B., & Peterson, S. E. (1986). The sentence verification technique as a measure of listening and reading comprehension. Educational and Psychological Research, 6, 299-314.

Royer, J. M., Lynch, D. J., Hambleton, R. K., & Bulgarelli, C. (1984). Using the sentence verification technique to assess the comprehension of technical text as a function of subject matter expertise. American Educational Research Journal, 21, 839-869.

Royer, J. M., Marchant, H., Sinatra, G., & Lovejoy, D. (in submission). The prediction of college course performance from reading comprehension performance: Evidence for general and specific factors.

Royer, J. M., Tirre, W. C., Sinatra, G. M. and Greene, B. A. (in press). The assessment of on-line computer-presented text. Journal of Educational Research.

Saxe, G.B. (1985). Effects of schooling on arithmetical

understanding: Studies with Oksapmin children in Papua New Guinea. Journal of Educational Psychology, 77, 503-513.

Schwartz, A.J. (1971). A comparative study of values and achievement: Mexican-American and Anglo youth. Sociology of Education, 44, 438-462.

Spanos, G., Rhodes, N.C., Dale, T.C. & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. Cocking & J. Mestre (Eds.), Linguistic and Cultural Influences on Learning Mathematics (pp. 221-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. Psychological Review, 68, 301-340.

Taylor, O.L. (1978). Language issues and testing. Journal of Non White Concerns, 6, 125-132.

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Voss, J.F., Vesonder, G.T., & Spilich, G.J. (1980). Text generation and recall by high knowledge and low knowledge individuals. Journal of Verbal Learning and Verbal Behavior, 19, 651-667.

Walczyk, J. J., & Royer, J. M. (in submission). A program for constructing SVT tests: An alternative way of assessing text comprehension.

Wilen, D.K. & Sweeting, C.V.M. (1986). Assessment of limited English proficient students. School Psychology Review, 15,

59-75.

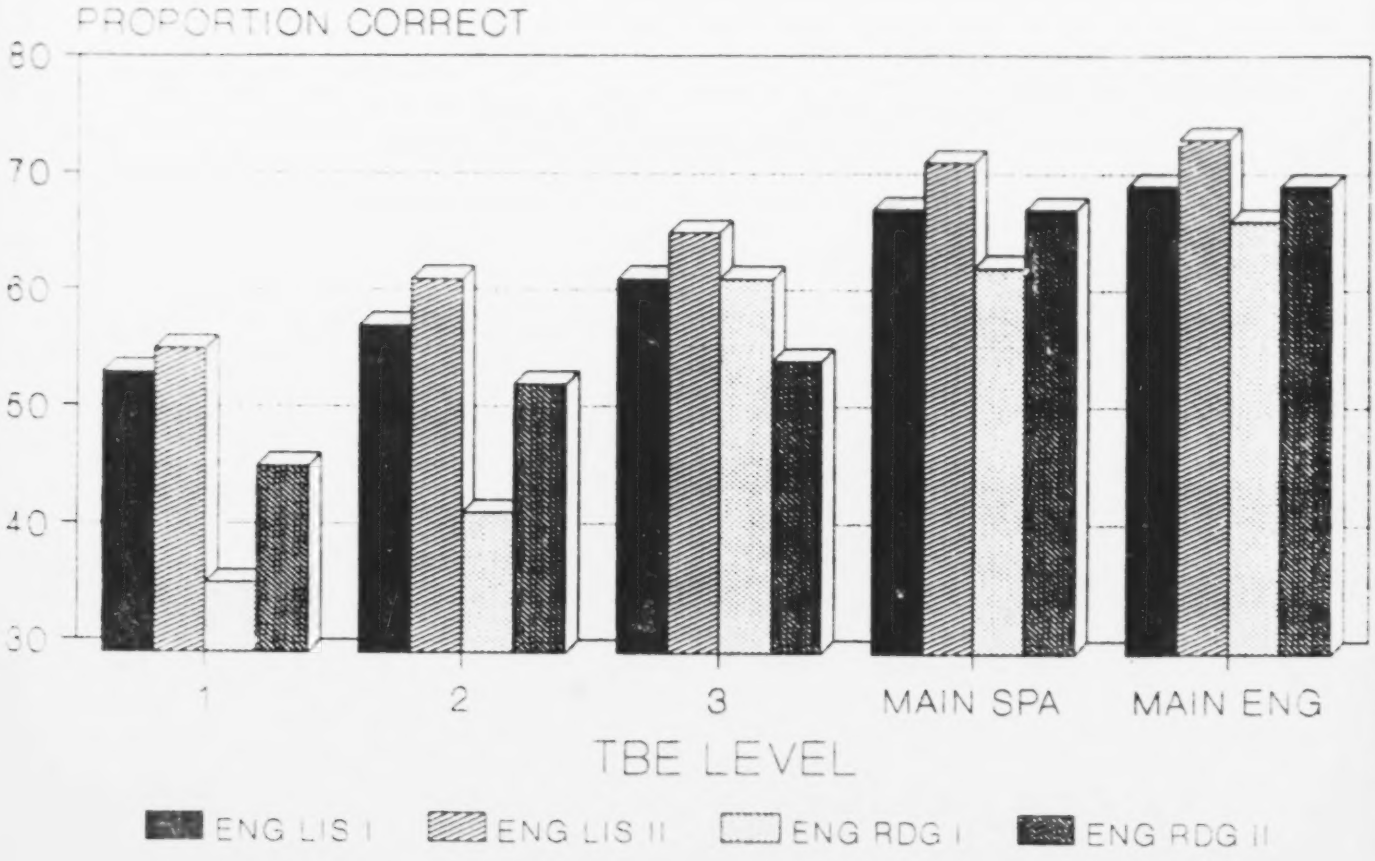
Table 1

Correlations Between Spanish and English Listening and Reading Performance on Three Different Test Occasions

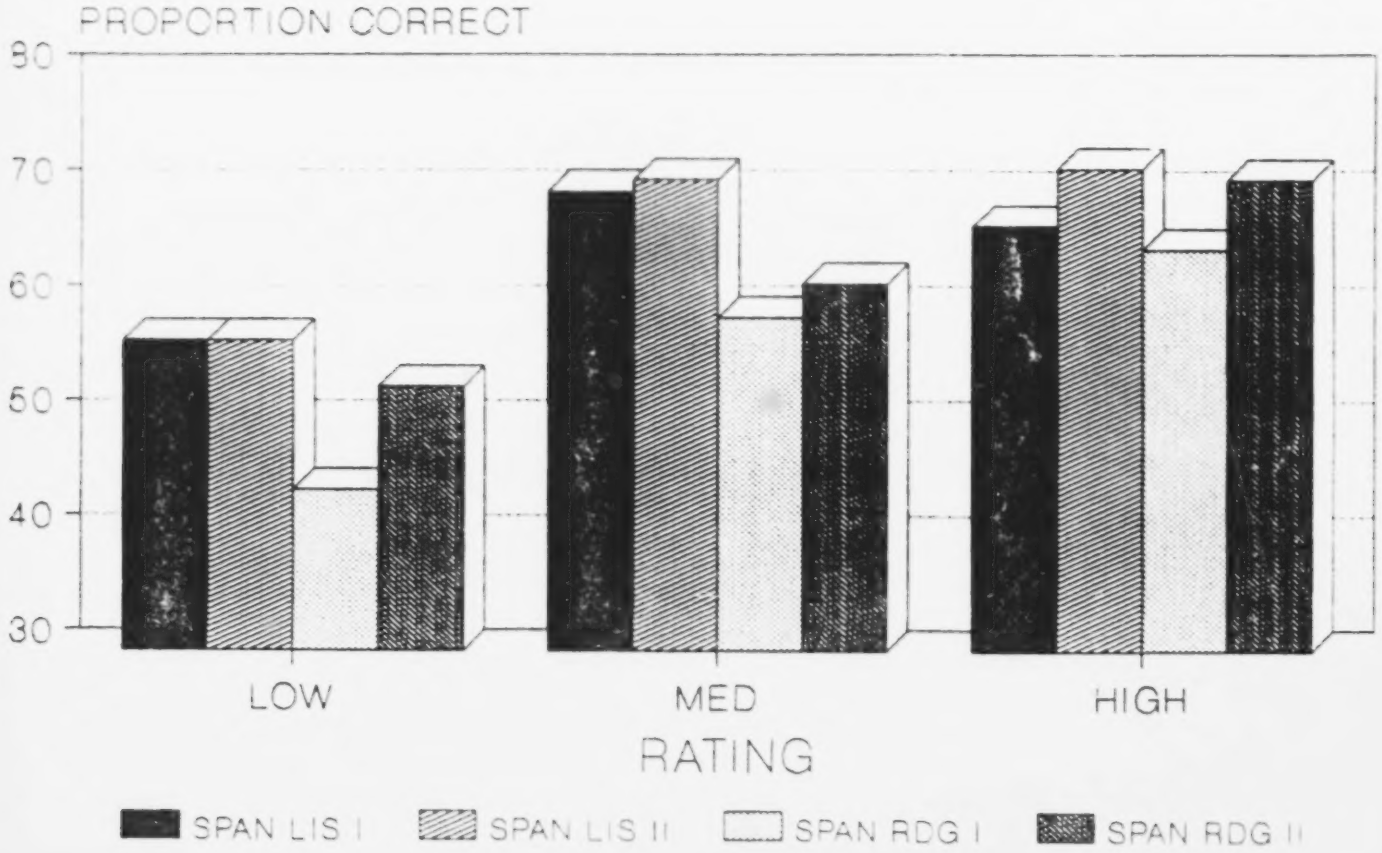
Variable	Correlations											
	1	2	3	4	5	6	7	8	9	10	11	12
1 Span Lis 1	1.0											
2 Span Rdg 1	.07	1.0										
3 Span Lis 2	.14	.08	1.0									
4 Span Rdg 2	.21	.16	.46*	1.0								
5 Span Lis 3	.10	.28	.40*	.15	1.0							
6 Span Rdg 3	.32	.14	.00	.35*	.32*	1.0						
7 Eng Lis 1	-.10	.40*	.02	.05	-.19	-.14	1.0					
8 Eng Rdg 1	-.17	-.19	.08	-.16	-.18	.10	.14	1.0				
9 Eng Lis 2	.07	-.22	.29	.26	.15	-.01	.09	-.01	1.0			
10 Eng Rdg 2	.01	.24	-.07	-.11	.01	.08	.11	-.02	.23	1.0		
11 Eng Lis 3	-.01	.16	.20	-.07	.05	-.17	.24	.16	.38*	.19	1.0	
12 Eng Rdg 3	.08	.12	.07	.29*	-.08	.18	.19	.19	.25	.08	.17	1.0

*p equal to or < .05

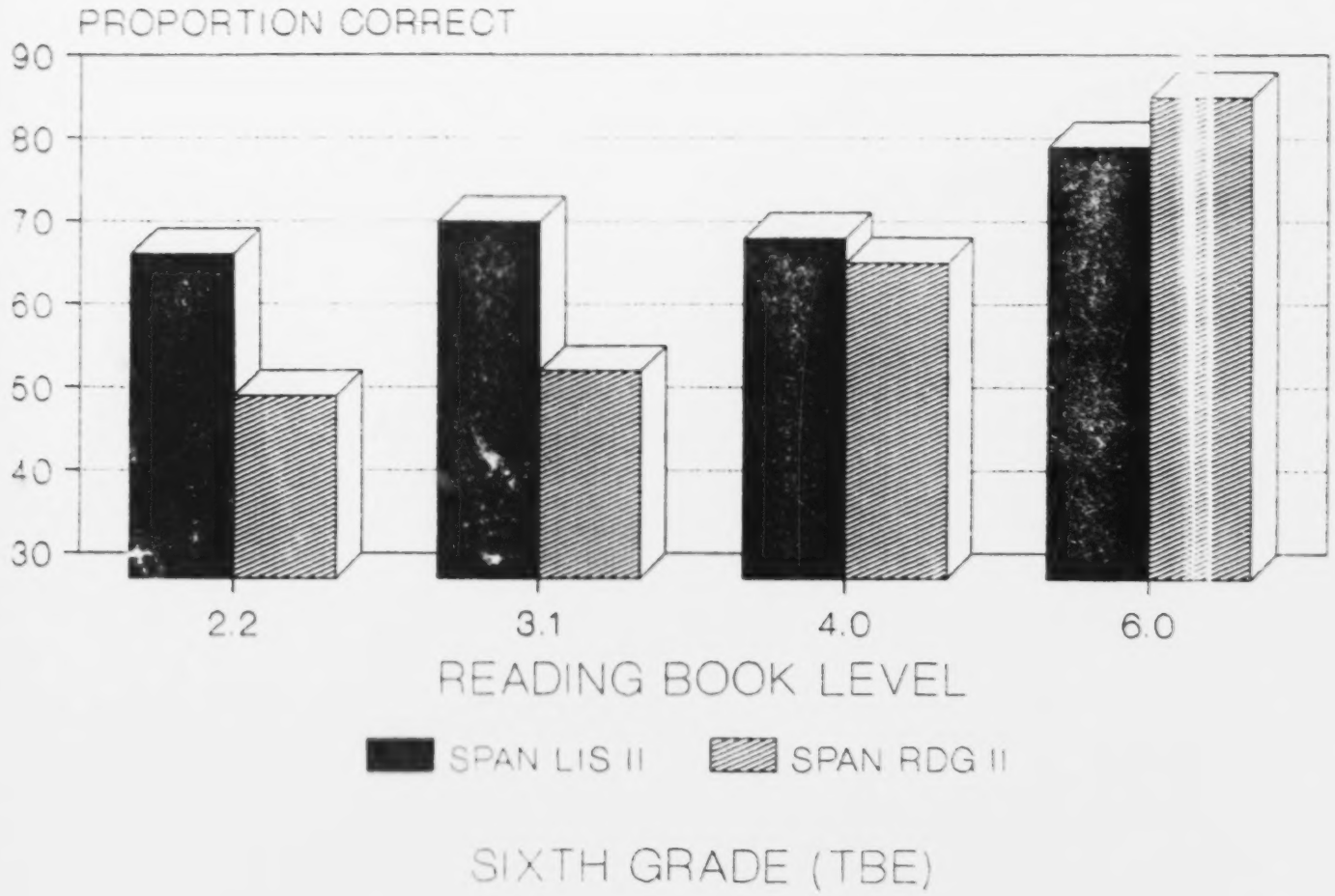
AVERAGE PERFORMANCE ON ENGLISH SVT TESTS AS A FUNCTION OF TBE LEVEL



SPANISH SVT PERFORMANCE AS A FUNCTION OF TEACHER RATINGS



AVERAGE PERFORMANCE ON SPANISH LIST AND RDG. SVT TESTS AS A FUNCTION OF READING BOOK LEVEL (2ND TEST ADMIN



SIXTH GRADE (TBE)